

## University of Groningen

### All agreed

Romeijn, Jan-Willem; Roy, Olivier

*Published in:*  
Theory and Decision

*DOI:*  
[10.1007/s11238-017-9633-9](https://doi.org/10.1007/s11238-017-9633-9)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Romeijn, J-W., & Roy, O. (2018). All agreed: Aumann meets DeGroot. *Theory and Decision*, 85(1), 41-60.  
<https://doi.org/10.1007/s11238-017-9633-9>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# All agreed: Aumann meets DeGroot

Jan-Willem Romeijn<sup>1</sup> · Olivier Roy<sup>2</sup>

Published online: 30 October 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** We represent consensus formation processes based on iterated opinion pooling (DeGroot in *J Am Stat Assoc* 69:118–121, 1974; Lehrer and Wagner in *Rational consensus in science and society: a philosophical and mathematical study*. Springer, Berlin, 1981) as a dynamic approach to common knowledge of posteriors (Aumann in *Ann Stat* 4(6):1236–1239, 1976; Geanakoplos and Polemarchakis in *J Econ Theory* 28:192–200, 1982). We thus provide a concrete and plausible Bayesian rationalization of consensus through iterated pooling. The link clarifies the conditions under which iterated pooling can be rationalized from a Bayesian perspective, and offers an understanding of iterated pooling in terms of higher-order beliefs.

**Keywords** Iterated pooling · Dynamic agreement theorem · Social epistemology

## 1 Introduction

This paper is about two influential models of convergence of probabilistic opinions: consensus through opinion pooling (DeGroot 1974; Lehrer and Wagner 1981), and agreement through Bayesian updates (Aumann 1976; Geanakoplos and Polemarchakis 1982).

---

✉ Jan-Willem Romeijn  
j.w.romeijn@rug.nl

Olivier Roy  
Olivier.Roy@uni-bayreuth.de

<sup>1</sup> Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

<sup>2</sup> Uni Bayreuth, Bayreuth, Germany

In linear pooling, upon learning what others believe each agent forms her new belief by taking a linear combination of the opinions of herself and of others, weighted by how much she trusts or respects them. By iterating this process sufficiently often the agents will converge to a fixed point in the space of opinions. In Bayesian models of agreement, on the other hand, upon learning what the others believe, the agents update their beliefs by Bayesian conditioning. Under the assumption of a common starting point and under the further assumption that agents know the type of information, but not the information content, that others have been given in the meantime, iteratively announcing the posteriors will lead the agents to agree.

While it has been claimed that both models offer a method to achieve consensus that is rational in its own right (Aumann 1976; Lehrer and Wagner 1981), doubts have been cast on the conceptual compatibility of pooling and Bayesian updating—see in particular the discussion in Bradley (2007) and Steele (2012). Moreover, Bayesian updating is often taken as the “basic normative standard” for rational belief change (cf. Bradley 2007, p. 12). This leads to the question whether iterated pooling can be given a Bayesian rationalisation. Interestingly, this is what Aumann writes in his seminal paper:

It seems to me that the Harsanyi doctrine is implicit in much of [the literature on opinion pooling]...The result of this paper may be considered a theoretical foundation for the reconciliation of subjective probabilities [i.e., by means of pooling].

To the best of our knowledge, Aumann’s suggestion was never converted into a formal result. Can we indeed reconstruct iterated pooling in Bayesian terms, in such a way that the sequence of steps towards consensus taken by iterated pooling can be motivated and accounted for?

Several results have already shown necessary and sufficient conditions for the Bayesian representation of a single pooling operation (Genest and Schervish 1985; Bonnay and Cozic 2015; omitted for blind review, 2015). So there is a Bayesian representation of agents engaged in pooling, and hence there is, in a strict sense, a formal account of iterated pooling as well. But what is lacking is an account of why these iteration steps should be taken in the first place. What underpins the specific social influence that is modeled by iterated pooling? The idea behind pooling is that the agents trust or respect, to various degree, the opinion of the others. But there is no such thing as trust, at least on the face of it, in the Bayesian redescription of pooling, and the iterations of pooling lack proper motivation.

We provide a representation theorem that makes good on Aumann’s suggestion and thereby answers the above questions: consensus via pooling can be represented as an Aumann-style agreement through updating. Iterated pooling is thus rationalized according to Bayesian standards, as a matter of higher-order reasoning. More precisely, the repeated adjustments of pooling can be represented as steps in which the agents not only rely on their own beliefs regarding the reliability of the others, but also on their beliefs regarding the others’ beliefs in their reliability, and so on. The representation offers a concrete interpretation of the epistemology of iterated pooling: pooling steps constitute an exchange of information, and the trust that agents put in each other

translates into beliefs regarding the reliability of others on different levels of the belief hierarchy.

In what follows we assume familiarity with iterated pooling, and so only introduce it briefly in Sect. 3. Agreement theorems and their dynamic versions, which underlie our representation, are far less discussed in the social epistemology literature. So we present them in some detail in Sect. 2. We prove the main theorem in Sect. 3. Section 4 discusses the interpretation of pooling that the representation gives, and considers the Bayesian model as an expression of conditions of applicability.

## 2 Dynamic agreement result

We survey here the original agreement theorem of [Aumann \(1976\)](#), and the dynamic formulation of the result by [Geanakoplos and Polemarchakis \(1982\)](#). We end with a detailed analysis of the events on which agents update in their dynamic approach to common knowledge. This is crucial for the proof of the main theorem in the next section.

### 2.1 The agreement theorem

First we define the algebraic structure that will be used in what follows. Let  $W_A$  and  $W_E$  be finite sets with elements  $w_A$  and  $w_E$ , respectively, and let  $W = W_A \times W_E$  be a finite set of possible worlds.<sup>1</sup> We can partition  $W$  into sets of worlds  $[w_A] = \{w : w = \langle w_A, v \rangle \text{ with } v \in W_E\}$ , and similarly into sets of worlds  $[w_E] = \{w : w = \langle v, w_E \rangle \text{ with } v \in W_A\}$ . We denote the set of all such sets  $[w_A]$  with  $[W_A]$  and similarly for  $[W_E]$ . Now we define  $\mathcal{A}$  as the algebra generated by the set  $[W_A]$ , that is, the algebra of factual propositions  $\mathcal{A} = \mathcal{P}(W_A) \times \{W_E\}$ , and similarly we define the algebra of epistemic propositions  $\mathcal{E} = \mathcal{P}(W_E) \times \{W_A\}$ . Finally, we write  $\mathcal{U} = \mathcal{P}(W)$  for the algebra of all propositions.

We now give more detail on the interpretation of these algebras. Consider two agents, Raquel and Quassim. The elements  $A \in \mathcal{A}$  represent material facts. Next to this, we have epistemic facts, expressed as elements of  $\mathcal{E}$ . These concern the opinions of Raquel and Quassim about propositions  $A$  but also the opinions that the two agents have about each others opinions, and so on. Every world  $w$  is characterized by a truth valuation over all the material facts in  $\mathcal{A}$  and over all the epistemic facts about the agents in  $\mathcal{E}$ .

The opinions of Raquel and Quassim are expressed by probability functions  $P_R$  and  $P_Q$ , both of which are defined on the algebra  $\mathcal{U}$ . Propositions about opinions will be denoted by means of upper corners,  $\lceil \cdot \rceil$ . For example, the proposition that Quassim has the opinion  $P_Q(A) = q$  is written as  $\lceil P_Q(A) = q \rceil$ . In what follows we only consider opinions about  $A$  and, therefore, abbreviate this to  $\lceil q \rceil$ , and similarly for Raquel, writing  $\lceil P_R(A) = r \rceil$  as  $\lceil r \rceil$ . Raquel might assign a probability to the proposition  $\lceil q \rceil$ , expressed in  $P_R(\lceil q \rceil)$ , and similarly Quassim may have an

<sup>1</sup> As will be explained below, we restrict attention to a finite approximation of consensus by iterated pooling. So it will be sufficient to work with finite algebras in the Bayesian models as well.

opinion about the proposition that Raquel assigns some probability  $p$  to his opinion,  $P_Q(\ulcorner P_R(\ulcorner q \urcorner) = p \urcorner)$ .

Such propositions are included in the algebra  $\mathcal{E}$  of epistemic facts about Raquel and Quassim. Recall that the set of worlds  $W$  and the algebras defined on it are all finite. This also holds for  $\mathcal{E}$ . Hence, we do not consider all possible probability assignments for the agents. We only consider the probability assignments and corresponding propositions salient for present purposes, namely those that are needed for the representation of iterated pooling. We will elaborate on these propositions below but first we introduce further elements of the algebra  $\mathcal{E}$ .

Quassim and Raquel may also have knowledge, rather than mere opinions, about each others' opinions.<sup>2</sup> For instance, Raquel may know that  $P_Q(A) = q$ , written as  $K_R \ulcorner P_Q(A) = q \urcorner$ . With such propositions on knowledge or information states of Raquel and Quassim, we now associate the further algebras  $\mathcal{R}$  and  $\mathcal{Q}$ . These algebras are based on specific partitions of  $W$ , the so-called information partitions. The members of these partitions,  $R_j$  and  $Q_k$ , represent the maximally specific information states in which Raquel and Quassim might find themselves, i.e., the maximally specific propositions that they might know.

We can now specify the knowledge operators  $K_R$  and  $K_Q$ , and the notion of common knowledge, in terms of these information partitions. We write  $R_{I(w)}$  for the partition cell  $R_j$  that contains  $w$ , i.e.,  $I(w)$  is a function that returns the value for  $j$  such that  $w \in R_j$ . The proposition that Raquel knows some proposition  $U$ , written  $\ulcorner K_R U \urcorner$ , is the set of all worlds that are a member of a partition cell fully included in  $U$ :

$$\ulcorner K_R U \urcorner = \{w : R_{I(w)} \subseteq U\}.$$

Accordingly, Quassim knows that Raquel knows  $U$  in all the worlds whose partition cell  $Q_{I(w)}$  is fully included in  $K_R U$ . Worlds included in a partition cell  $Q_k$  that overlaps with, but is not fully included in  $K_R U$  do not qualify. Formally:

$$\ulcorner K_Q \ulcorner K_R U \urcorner \urcorner = \{w : Q_{I(w)} \subseteq \ulcorner K_R U \urcorner\}.$$

The entire hierarchy of higher-order knowledge propositions can be built up in this way. We define “everybody knows  $U$  up to level  $n$ ” inductively as follows:  $E^0 U$  is just  $U$ , and  $E^{n+1} U$  is  $\ulcorner K_R \ulcorner E^n U \urcorner \urcorner \cap \ulcorner K_Q \ulcorner E^n U \urcorner \urcorner$ . Following the standard definition (Aumann 1976), the proposition that  $U$  is common knowledge between Raquel and Quassim, written  $C_{RQ} U$ , is obtained by taking the limit of  $E^n U$  for increasing  $n$ :  $C_{RQ} U = \bigcap_{n < \omega} E^n U$ . To say that  $U$  is common knowledge for Raquel and Quassim is to say that we cannot, by some series of mutually overlapping  $R_j$ 's and  $Q_k$ 's, reach a world outside of  $U$ : the set  $U$  is closed under the application of knowledge operators.

<sup>2</sup> The proposition that Raquel knows Quassim's opinion is logically stronger than the proposition  $\ulcorner P_R(\ulcorner P_Q(A) = q \urcorner) = 1 \urcorner$ . Judging some proposition to have unit probability is not enough for saying that it will obtain. For example, if one samples individuals from an infinite population in which there is only one individual satisfying some property, then one will assign the satisfaction of that property for the next individual a unit probability even while, strictly speaking, one does not know that the next individual will satisfy it.

To express the agreement theorem, we must suppose that the opinions of Raquel and Quassim  $P_R$  and  $P_Q$  have been obtained from updating a regular common prior  $P$ , i.e., a prior that is shared by both agents and non-zero everywhere on  $\mathcal{U}$ , on private information,  $R_0$  and  $Q_0$ , respectively:

$$P_R(\cdot) = P(\cdot|R_0), \quad (1)$$

$$P_Q(\cdot) = P(\cdot|Q_0). \quad (2)$$

Clearly, Raquel and Quassim do not know what private information the other has. The information partitions, however, are common knowledge. Raquel and Quassim know that the other agent receives information in the form of a proposition from  $\mathcal{Q}$  and  $\mathcal{R}$ , respectively, and they know what these algebras look like.

### Agreement Theorem (Aumann 1976)

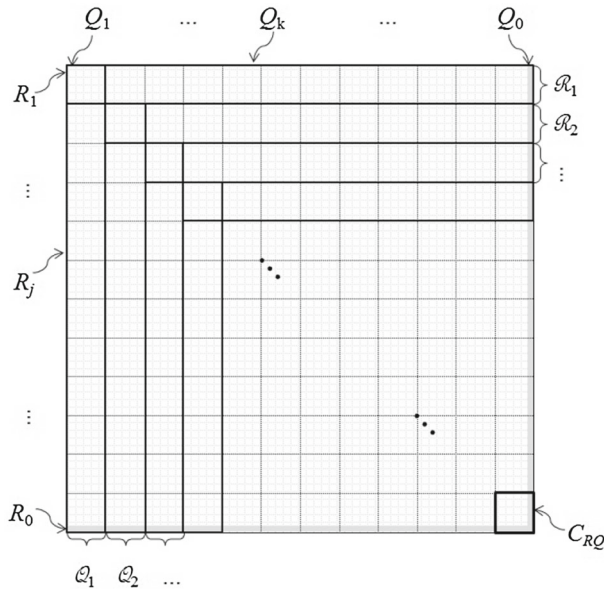
Assume that Raquel and Quassim have a common prior and that in a world  $w$  their posterior beliefs  $P_R(A) = r$  and  $P_Q(A) = q$  are common knowledge, viz.  $w \in C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)$ . Then, their posterior beliefs coincide in that world,  $r = q$ .

This result rests on the properties of the set of worlds  $C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)$ , i.e., worlds in which common knowledge of the posteriors obtains. In particular, note that for any element  $R_j$  from Raquel's information partition we must have  $P(A|R_j \cap C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)) = r$  because if not, Raquel or Quassim would consider it possible that Raquel's opinion were unequal to  $r$ , thus violating common knowledge. By the law of total probability, we therefore have  $P(A|C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)) = r$ . The same will hold for Quassim, i.e.,  $P(A|Q_k \cap C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)) = q$  for any  $k$  and hence  $P(A|C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)) = q$ , and thus  $r = q$ .

## 2.2 Dynamic agreement

The original agreement theorem is static. It establishes the equality of the posteriors under the assumption that common knowledge has obtained. Our Bayesian representation of the pooling model of consensus rather rests on a dynamic version of the agreement theorem due to [Geanakoplos and Polemarchakis \(1982\)](#). They show iterated exchange of opinions will lead to common knowledge among Raquel and Quassim and hence, under common prior, to agreement. Importantly, there are only finitely many propositions on opinions that are exchanged in the approach to common knowledge detailed below, and this explains that the algebras  $\mathcal{R}$  and  $\mathcal{Q}$  are indeed finite.

For the dynamic approach to common knowledge, we again assume a common prior  $P$  over  $\mathcal{U}$ , so that Raquel and Quassim start with the same opinion,  $r_0 = P(A) = q_0$ , with  $\ulcorner r_0 \urcorner = W = \ulcorner q_0 \urcorner$ . Our agents acquire information privately,  $R_0$  and  $Q_0$ , update on this information and so arrive at the starting opinions of Eqs. (1) and (2). Raquel and Quassim then disclose their opinions about  $A$  and, assuming common knowledge of the information partitions, they each update on this information, revising their opinion on  $A$  according to what the other person thinks, and according to what they now know the other person knows about them. After this the cycle repeats. Importantly, they



**Fig. 1** The Aumann structure for Raquel and Quassim with common knowledge of posteriors showing as the proposition in bold

update by conditioning so that their subsequent probability assignments can all be traced back to a common prior  $P$ . Raquel and Quassim go through belief states that are conditioned on ever more specific propositions.

Focusing first on Raquel's epistemic development, at any stage she updates by conditioning on the opinion that Quassim holds at that stage. For instance, when Raquel learns Quassim's initial opinion,  $P_Q(A) = q_1$ , she conditions her probability assignment according to  $P_R(A | \ulcorner q_1 \urcorner) = r_2$ . In particular, she will progressively eliminate some set of elements  $Q_k$  of Quassim's information partition that cannot be squared with Quassim's current opinion. At every stage, her reasoning is: "If Quassim had initially learnt the information  $Q_k$ , he would not now have the opinion  $q_i$  about  $A$ ". In this manner, Raquel iteratively reconstructs the element  $Q_0$  that Quassim must have learnt, up to equivalence in terms of Quassim's posterior. The same holds for Quassim, who eliminates elements from the information partition of Raquel up to equivalence with  $R_0$ , at which point Raquel's opinion does not allow him to eliminate any further elements. This is when common knowledge of posteriors is reached.

Figure 1 illustrates this iterative process for Raquel and Quassim. The cells are sets of possible worlds. The thinly delineated rows and columns represent elements  $R_j$  and  $Q_k$ , respectively. Raquel and Quassim learn that the actual world is in the last row  $R_0$  and in the rightmost column  $Q_0$ , respectively, both of which are coloured grey. Accordingly, the actual world is in the bottom right grey corner. The coarse-grained rows  $R_i$  represent sets of elements  $R_j$  that Quassim eliminates at subsequent update stages, after learning Raquel's subsequent opinions  $r_i$ , and vice versa for the columns  $Q_i$ , which consist of elements  $Q_k$ . The common knowledge proposition  $C_{RQ}(\ulcorner r \urcorner \cap \ulcorner q \urcorner)$  is delineated in bold.

At each round, Quassim and Raquel do not only learn the opinion of the other, though. They also learn that the other has learned their own previous opinion. The exchange of opinion is completely transparent for both of them. So Quassim will himself alter the set of worlds that he takes to be accessible for Raquel: he eliminates all the worlds outside  $K_R \lceil q_0 \rceil$  from Raquel's domain, and similarly for Raquel, who eliminates all worlds outside  $K_Q \lceil r_0 \rceil$ .

It is important to observe that even though this exchange of opinion is completely public between Raquel and Quassim, a single round of simultaneous updates on each other's opinions needs not to create common knowledge of their posteriors. It does create common knowledge of what these posteriors were *before* the announcement. But upon learning the opinion of the other about  $A$ , both Raquel and Quassim can change their own opinion about  $A$ . When Quassim announces his current posterior to Raquel, he of course knows that Raquel has learned his opinion. But he also realizes that Raquel's old opinion  $r_1$  may not be her new opinion anymore. She could have changed her mind upon learning Quassim's opinion, and he needs not to know exactly how. The same goes for Raquel. So even though a simultaneous, fully public exchange of posteriors between Raquel and Quassim creates common knowledge of what these posteriors were before the exchange, it needs not to create common knowledge of what the posteriors are after. To reach that they need more steps.

Updates are repeated every time Quassim and Raquel disclose their new opinions, creating two sequences of events,  $\lceil r_i \rceil$  and  $\lceil q_i \rceil$ . At each step, these sets cover fewer elements of the information partitions  $\mathcal{R}$  and  $\mathcal{Q}$ , respectively. The sets  $\lceil r_i \rceil$  and  $\lceil q_i \rceil$  hence form so-called filtrations, i.e., sets with index  $i + 1$  are nested in those with index  $i$ . Referring back to the diagrammatic representation above, we partition  $W$  according to these filtrations into:

$$\mathcal{R}_i = \lceil r_{i-1} \rceil \setminus \lceil r_i \rceil, \quad (3)$$

$$\mathcal{Q}_i = \lceil q_{i-1} \rceil \setminus \lceil q_i \rceil. \quad (4)$$

So the sets  $\mathcal{R}_i$  and  $\mathcal{Q}_i$  consist of those elements of the information partitions  $\mathcal{R}_j$  and  $\mathcal{Q}_k$  that are ruled out in update stage  $i$ . As indicated, they are coarse-grainings of the information partitions  $\mathcal{R}$  and  $\mathcal{Q}$ . Notably, they do not coincide with the sets of worlds eliminated at every update stage because Raquel and Quassim have already learnt  $R_0$  and  $Q_0$ . The agents eliminate elements of the information partition of each other only insofar as these overlap with  $R_0$  and  $Q_0$ , respectively, and they record the elimination of elements of their own information partitions only to keep track of what the other is thinking of them.

We are now ready to offer a recursive definition of the events  $\lceil r_i \rceil$  and  $\lceil q_i \rceil$ , and hence of the sets  $\mathcal{R}_i$  and  $\mathcal{Q}_i$ , in terms of specific probability assignments within these events:

$$\begin{aligned} \lceil q_i \rceil &= \{\cup_k Q_k : P(A|Q_k \cap \lceil r_{i-1} \rceil) = q_i\} \cap \lceil q_{i-1} \rceil, \\ \lceil r_i \rceil &= \{\cup_j R_j : P(A|R_j \cap \lceil q_{i-1} \rceil) = r_i\} \cap \lceil r_{i-1} \rceil. \end{aligned}$$

Notice that the events  $\lceil q_i \rceil$  and  $\lceil r_i \rceil$  include all the  $Q_k$  and  $R_j$  that are consistent with the information that Raquel and Quassim, respectively, have about the opinion states



of the other. Equations (3) and (4) show how these events relate to the events that Raquel and Quassim eliminate at every stage.

In the definition of the events  $\lceil r_i \rceil$  and  $\lceil q_i \rceil$ , we can see the interplay between Raquel and Quassim at work. When Raquel wants to determine which elements  $Q_k$  are included in  $\lceil q_i \rceil$ , she restricts attention to those  $Q_k$  inside  $\lceil q_{i-1} \rceil$ , which have not been ruled out. But she then looks whether  $P(A|Q_k \cap \lceil r_{i-1} \rceil) = q_i$ , because she wants to see if Quassim, who now has the opinion  $q_i$  and who has updated on her earlier opinion  $\lceil r_{i-1} \rceil$ , could have learnt  $Q_k$ . To do this, however, she needs to know precisely what Quassim took away from her earlier opinion,  $\lceil r_{i-1} \rceil$ , and for that she needs the definitions of both  $\lceil r_{i-2} \rceil$  and  $\lceil q_{i-2} \rceil$ , and so on, all the way down the epistemic hierarchy.

With this in place, we can specify the two update series that govern the opinions of Raquel and Quassim:

$$\begin{aligned} r_{i+1} &= P_R(A|\lceil q_i \rceil) = P(A|R_0 \cap \lceil q_i \rceil), \\ q_{i+1} &= P_Q(A|\lceil r_i \rceil) = P(A|Q_0 \cap \lceil r_i \rceil). \end{aligned}$$

So these updates affect the probabilities that Raquel and Quassim have for  $A$  and lead them from  $r_i$  to  $r_{i+1}$ , and from  $q_i$  to  $q_{i+1}$ , starting from  $r_1$  and  $q_1$ , respectively. In addition, they both update with the knowledge that the other agent learnt something. However, this second update has no impact on their probability for  $A$ :

$$\begin{aligned} P_R(A|\lceil q_i \rceil) &= P_R(A|\lceil q_i \rceil \cap K_Q \lceil r_i \rceil), \\ P_Q(A|\lceil r_i \rceil) &= P_Q(A|\lceil r_i \rceil \cap K_R \lceil q_i \rceil). \end{aligned}$$

The update is nevertheless important for our understanding of the approach to common knowledge. After having updated on the other agent knowing their opinion, they do not know anymore what the other thinks about  $A$ . So this update explains that Raquel is ready to accommodate Quassim's new opinion  $q_{i+1}$ , and that Quassim is ready for Raquel's new opinion  $r_{i+1}$  too, after both have gone through update round  $i$ .

The exchange of opinions will continue until the opinions of Quassim and Raquel concur at some stage,  $q_n = r_n$  say. To see how this obtains, recall that at stage  $n$  Raquel only considers worlds within  $R_0 \cap \lceil q_{n-1} \rceil$  possible and, similarly, Quassim is left only with worlds in  $Q_0 \cap \lceil r_{n-1} \rceil$ . Moreover, Raquel and Quassim know of each other that they only consider worlds within  $\lceil q_{n-1} \rceil \cap \lceil r_{n-1} \rceil$  possible. But imagine that for every  $Q_k \in \lceil q_{n-1} \rceil$  we have that  $P(A|Q_k \cap \lceil r_{n-1} \rceil) = q_n$ , and that for every  $R_k \in \lceil r_{n-1} \rceil$  we have that  $P(A|R_k \cap \lceil q_{n-1} \rceil) = r_n$ . In that case, the only probability for  $A$  that Raquel might have is  $r_n$ , and similarly  $q_n$  is the only probability left open for Quassim. At any earlier stage, both will have had some variation in the probability of  $A$  over the elements from their information partitions within  $\lceil q_{n-1} \rceil \cap \lceil r_{n-1} \rceil$ , but at round  $n$  this probability is constant for both.

This is the point of the dynamic version of the agreement theorem due to [Geanakoplos and Polemarchakis \(1982\)](#), i.e., the result that “we cannot disagree forever”. For an informal grasp of their result, we observe that at round  $n$  the opinions of Raquel and Quassim are indeed common knowledge: neither can conceive that the other will have a different probability for  $A$  than what they have, and they also know this of

each other, know that they know this, and so on. In terms of the characterisation of common knowledge above: there is no series of mutually overlapping  $R_j$ 's and  $Q_k$ 's such that either agent can arrive at a world at which the probability assignment for  $A$  differs from  $r_n$  or  $q_n$ , because these assignments are constant over all accessible rows and columns. This fact, which establishes the common knowledge of the probability assignments, namely that the probabilities are constant over all remaining sets within both information partitions, is also responsible for the equality of these assignments. We obtain full agreement about  $A$ ,  $r_n = q_n$ , by the marginalization argument already given in the context of Aumann's static result.

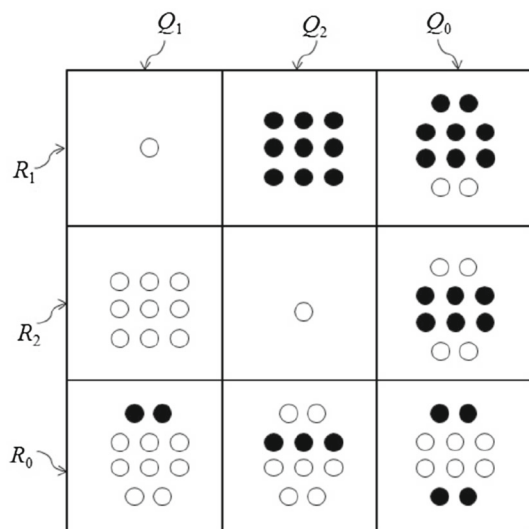
### 2.3 Approach to common knowledge: example

It will be insightful to provide a brief example of the approach to common knowledge, with which we can illustrate the diagrammatic representation given above. The example is inspired by the one provided in [Geanakoplos and Polemarchakis \(1982\)](#) but it is constructed to match a simple pooling case, as presented in the next section.

Consider the set of worlds and the information partitions given in Fig. 2. The dots represent equiprobable possible worlds. Solid dots are worlds in which  $A$  is true, open dots represent worlds at which  $\neg A$  is true. At the outset Raquel learns  $R_0$  and so she has a probability of  $r_1 = P(A|R_0) = \frac{9}{30} = \frac{3}{10}$ , while Quassim learns  $Q_0$  and so has  $q_1 = P(A|Q_0) = \frac{18}{30} = \frac{3}{5}$ .

Raquel and Quassim then tell each other their probability assignments. Upon hearing  $q_1$ , Raquel concludes that Quassim cannot have learnt  $Q_1$  because if he had, he would have given  $A$  a probability of  $\frac{1}{10}$  instead. But he could still have learnt  $Q_2$  because  $P(A|Q_2) = \frac{12}{20} = \frac{3}{5}$  as well. Hence, Raquel takes  $\lceil q_1 \rceil = Q_0 \cup Q_2$  and eliminates  $Q_1$ . Her new probability for  $A$  becomes  $r_2 = P(A|R_0 \cap \lceil q_1 \rceil) = \frac{7}{20}$ . Upon

**Fig. 2** Example of an Aumann structure for Raquel and Quassim



hearing  $r_1$ , Quassim in his turn concludes that Raquel can still have learnt  $R_2$  but that she cannot have learnt  $R_1$  because if she had, she would have given  $A$  a probability of  $\frac{17}{20}$ . Hence,  $\lceil r_1 \rceil = R_0 \cup R_2$ , so that Quassim's new probability for  $A$  becomes  $q_2 = P(A|Q_0 \cap \lceil r_1 \rceil) = \frac{10}{20} = \frac{1}{2}$ .

Notice that Raquel and Quassim both realize that the other agent will have eliminated some worlds, namely from the sets  $Q_1$  and  $R_1$ , respectively. Because they do not know what information the other obtained in the first place, they cannot derive from that what the new probability assignment of the other agent will be. Raquel and Quassim really learn something new when tell each other their new probability assignments,  $r_2 = \frac{7}{20}$  and  $q_2 = \frac{1}{2}$ .

Upon hearing  $q_2$ , Raquel concludes that Quassim cannot have learnt  $Q_2$  because if he had, he would have had  $q_2 = \frac{3}{11}$ . Importantly, to determine this Raquel uses her understanding of Quassim's epistemic situation, including his understanding of hers. She knows that he has updated on  $\lceil r_1 \rceil$ , i.e., eliminated  $R_1$ , and she evaluates what would have been Quassim's opinion if he had learnt  $Q_2$  at the outset and then accommodated  $\lceil r_1 \rceil$ , finding that he would then have given  $A$  a probability of  $\frac{3}{11}$ . She thus eliminates all of  $Q_2$ , taking  $\lceil q_2 \rceil = Q_0$ . Her new probability for  $A$  becomes  $r_3 = P(A|R_0 \cap \lceil q_2 \rceil) = \frac{4}{10} = \frac{2}{5}$ . Furthermore, upon hearing  $r_2$ , Quassim concludes that Raquel cannot have learnt  $R_2$  because if she had, she would have had  $r_2 = \frac{6}{11}$ , going through a similar train of higher-order thought as was just described for Raquel. He eliminates  $R_2$  altogether, taking  $\lceil r_2 \rceil = R_0$ , and his new probability for  $A$  becomes  $q_3 = P(A|Q_0 \cap \lceil r_2 \rceil) = \frac{2}{5}$  as well! He comes to agree with Raquel.

At this point, Raquel and Quassim have achieved common knowledge and so agree on the probability of  $A$ . In this simple case, agreement is reached because both agents derived what the other agent actually learnt at the outset. But this is not necessary for the agreement to obtain: it is enough if, among the remaining elements of the information partition, the probability of  $A$  among the worlds still accessible for the other agent is constant.

### 3 Representing opinion pooling

We turn to the target of the paper: to show that a process of iterated pooling among Raquel and Quassim can be framed as a dynamic approach to common knowledge, through repeated Bayesian updates by Raquel and Quassim on each other's posteriors. In this way, we establish that iterated pooling can be rationalized as a specific Bayesian procedure. We first describe a finite version of consensus through iterated pooling, which prepares for the proof of the main theorem. At the end of the section, we consider an extension and a refinement of the result.

#### 3.1 Consensus in finitely many steps

Consider Raquel and Quassim who both have an opinion about  $A$ , to wit,  $P_R^1(A) = r_1$  and  $P_Q^1(A) = q_1$ . Linear pooling determines that the subsequent opinion of Raquel  $P_R^2(A)$  is given by

$$r_2 = w_R q_1 + (1 - w_R) r_1. \quad (5)$$

The parameter  $w_R \in [0, 1]$  specifies to what extent the updated opinion of Raquel will move towards that of Quassim. By way of interpretation: it measures the trust that Raquel has in Quassim. With  $w_Q$  we can determine the subsequent opinion of Quassim in the same way.<sup>3</sup>

The pooling operation can be iterated with the same parameters, resulting in a series of opinions of both Raquel and Quassim:

$$\langle r_1, q_1 \rangle, \langle r_2, q_2 \rangle, \dots, \langle r_{n-1}, q_{n-1} \rangle, \langle r_n, q_n \rangle \dots \quad (6)$$

If the weights are both constant and sit within the open unit interval, the process is guaranteed to converge to consensus,  $\langle p, p \rangle$ . This condition is not necessary; see Jackson (2008, chap. 8) for a more elaborate presentation of the model. The consensus opinion is a fixed point in the Markov process for which the weights  $w_R$  and  $w_Q$  determine the transition matrix. Intuitively, Raquel and Quassim will traverse the distance between them in steps proportional to the weights they assign to each other, and so meet at an intermediate point.

For a numerical example, we can briefly return to Sect. 2.3. We fix  $r_1 = \frac{3}{10}$  and  $q_1 = \frac{3}{5}$  and we choose  $w_R = \frac{1}{6}$  and  $w_Q = \frac{1}{3}$ . We then find  $r_2 = \frac{7}{20}$  and  $q_2 = \frac{1}{2}$ , then  $r_3 = \frac{9}{24}$  and  $q_3 = \frac{9}{20}$ , and so on, until we reach  $p = r_1 + w_R(q_1 - r_1)/(w_R + w_Q) = \frac{2}{5}$  after infinitely many iterations.

Except when the weights are the same for all agents, iterated pooling with constant weights arrives at consensus only in the limit. But in order to align this process to an approach to common knowledge, it is convenient to convert it to a series with finite length. Notably, if there is indeed a consensus point  $p$ , we can always close off the series this by adapting the weights in the final step  $n$ , namely by scaling up the weights proportionally so that they add up to 1:

$$w'_R = \frac{w_R}{w_R + w_Q} = 1 - w'_Q.$$

The result of closing off in this way coincides with the result of an indefinite iteration of pooling. In the example, we obtain  $w'_R = \frac{1}{3}$  and  $w'_Q = \frac{2}{3}$ . If we close off at step  $n = 3$ , we then go from  $r_2 = \frac{7}{20}$  and  $q_2 = \frac{1}{2}$  straight to  $r'_3 = \frac{2}{5} = q'_3$ , thereby following the approach to common knowledge of Sect. 2.3. Importantly, a series can be closed off in this way at any step  $n$ , and for large  $n$  the remaining discrepancy between  $r_{n-1}$  and  $q_{n-1}$  will become vanishingly small.

Considering that common knowledge is attained by a sequence of Bayesian updates, it might seem crucial that the pooling operation has a Bayesian representation. This representation is indeed available. Genest and Schervish (1985) prove that we can always organize our likelihoods in such a way that the prior and posterior conform to the pooling operation. More precisely, we can choose the likelihoods such that any

<sup>3</sup> This model dates back at least to French (1956) and has been developed by Stone (1961), DeGroot (1974) and numerous others. For philosophers, the classical treatment is Lehrer and Wagner (1981). In what follows we restrict ourselves to the model presented in DeGroot (1974).

pair of non-extremal priors for Raquel and Quassim will lead to the pair of posteriors determined by pooling. This leaves some room for variation in the probability assignments. Specifically, while the expectations that Raquel and Quassim have of each other must be centered on their own opinion, the shape of the probability distributions over the opinions of the other agent can be chosen freely. We return to this result and its use for our purposes in Sect. 3.3.

### 3.2 A common prior for iterated pooling

We have already provided an example of how a closed-off pooling process can be accommodated in an Aumann structure: the numbers of the example of iterated pooling match those of the approach to common knowledge in Sect. 2.3. In what follows, we are given a closed-off iterated pooling process,

$$\langle r_1, q_1 \rangle, \langle r_2, q_2 \rangle, \dots, \langle r_{n-1}, q_{n-1} \rangle, \langle r'_n, q'_n \rangle,$$

such that  $r'_n = q'_n$ , and we show that this process can be mapped onto the approach to common knowledge in general.

Recall that the Aumann structure consisted in a set of possible worlds  $W$ , carved up by two information partitions,  $\mathcal{R}$  and  $\mathcal{Q}$ . We first prove a generic Lemma, which will be applied repeatedly in the proof of the main theorem.

**Lemma** *Let  $S$  be a set of worlds and let  $\{S_i\}_{0 \leq i \leq m}$  be a filtration, with non-empty  $S_{i+1} \subsetneq S_i$  and  $S_0 = S$ . Let  $\mathcal{S}$  be the algebra generated by this filtration and the set  $A$ , where  $A \cap S_i \neq \emptyset$  for all  $i$ . Furthermore, let  $\langle p_1, p_2, \dots, p_m \rangle$  be a sequence of length  $m$  with  $0 < p_i < 1$  and  $p_i \neq p_{i'}$  unless  $i = i'$ . Then, there is a probability function  $P$  over  $\mathcal{S}$  such that for all  $0 \leq i < m$*

$$P(A|S_i) = p_{i+1},$$

where

$$0 < P(S_{i+1}|S_i) < \text{Min}\left(\frac{p_i}{p_{i+1}}, \frac{1-p_i}{1-p_{i+1}}\right). \quad (7)$$

*Proof* We prove this for a fixed value  $m$  and by induction over  $i$ . For  $i = 0$  the case is trivial: we can simply fix  $P(A|S) = P(A|S_0) = p_1$ . For the inductive step, assume that we satisfy  $P(A|S_{i'}) = p_{i'+1}$  for all  $i' < i$ , where  $i < m$ . In particular, we have that  $P(A|S_{i-1}) = p_i$ . We can now add the constraint that  $P(A|S_i) = p_{i+1}$  and hence show that we can satisfy the constraints  $P(A|S_{i'}) = p_{i'+1}$  for all  $i' < i + 1$ . Writing  $x = P(S_i|S_{i-1})$  and  $\bar{p}_{i+1} = P(A|\{S_{i-1} \setminus S_i\})$ , we have

$$p_i = xp_{i+1} + (1-x)\bar{p}_{i+1}.$$

We can solve for  $x$  and  $\bar{p}_{i+1}$  to obtain a range of possible values:

$$x = \frac{p_i - \bar{p}_{i+1}}{p_{i+1} - \bar{p}_{i+1}},$$

where we choose  $\bar{p}_{i+1}$  such that  $0 < x < 1$ , e.g., if  $p_i < p_{i+1}$  then  $0 < x < \frac{p_i}{p_{i+1}}$ , and if  $p_i > p_{i+1}$  then  $0 < x < \frac{1-p_i}{1-p_{i+1}}$ , which establishes the constraint Eq. (7).  $\square$

The Lemma establishes that an update series can lead to any sequence of probability values for the proposition  $A$ . So it can also replicate the probability assignments resulting from an iterated pooling process.<sup>4</sup> The Bayesian rationalization of pooling mentioned above falls under this general header but is otherwise far more specific, leaving less room for variation in the common prior. We will return to the use of this more specific equality for the purpose of linking agreement and consensus towards the end of this section. First, we prove the main theorem, by applying the Lemma to all the elements  $\mathcal{R}_j$  and  $\mathcal{Q}_k$ .

**Theorem** *Let  $\langle r_i, q_i \rangle_{i \leq n}$  be a sequence of probability assignments for a proposition  $A$ , resulting from an iterated pooling that is closed off at  $n$ . Then, there is always a finite two-person Aumann structure  $\langle W, \mathcal{R}, \mathcal{Q} \rangle$ , a regular common prior  $P$  over it, and a world  $w \in W$ , such that the approach to common knowledge through the exchange of posteriors matches the sequence produced by closed off iterated pooling:*

$$P_R(A | \ulcorner q_i \urcorner) = r_{i+1} \quad \text{and} \quad P_Q(A | \ulcorner r_i \urcorner) = q_{i+1}.$$

for all  $i = 0, 1, \dots, n$ , where  $P_R$  and  $P_Q$  are conditioned on events  $\mathcal{R}$  and  $\mathcal{Q}$  that contain  $w$ .

*Proof* The match between dynamic agreement and consensus formation involves the existence of a finite Aumann structure that can accommodate the exchange and iterated pooling of opinions, and the existence of a common prior over this structure that replicates the opinions. On the first claim, we will be brief. It is clear from the foregoing that such a finite Aumann structure can always be constructed. It merely requires us to make the structure large enough. The proof thus focuses on the second claim, which is proved in two parts. First, we establish that the common prior can accommodate the constraints on the posteriors for Raquel and Quassim. Then, we establish that the remaining freedom in the common prior easily suffices to meet the constraints needed for adequate definitions of the update events of Raquel and Quassim.

To accommodate the posteriors of Raquel and Quassim, the common prior must at least conform to the following:

$$P(A | R_0 \cap \ulcorner q_i \urcorner) = r_{i+1}, \quad (8)$$

$$P(A | Q_0 \cap \ulcorner r_i \urcorner) = q_{i+1}. \quad (9)$$

We can establish that these constraints can be satisfied by a direct application of the Lemma. Focusing on Raquel, we substitute  $R_0$  for  $S$  and for all  $0 \leq i < m$  we substitute  $\ulcorner q_i \urcorner$  for  $S_i$  and  $r_{i+1}$  for  $p_{i+1}$ , using  $m = n - 1$ . The Lemma then states that any consensus formation process can be accommodated. The same holds for the

<sup>4</sup> Notice that we require that  $p_i \neq p_{i'}$  unless  $i = i'$ . If we allow  $P(S_i | S_{i-1}) = x = 1$  we can also accommodate their equality. But for a sequence of opinions resulting from iterated pooling they will always differ, because the pooling weights are assumed to be smaller than 1.

process of Quassim, with  $Q_0$  for  $S$  and the  $S_i$ 's and  $p_{i+1}$ 's replaced by  $\lceil r_{i+1} \rceil$ 's and  $q_i$ 's, respectively.

Now consider an element  $Q_k$  of Quassim's information partition that Raquel eliminates when she updates on  $\lceil q_m \rceil$ , so  $Q_k \subset Q_m$ . Since  $Q_k \subset \lceil q_i \rceil$  for all  $i < m$ , the probability assignment within  $Q_k$  must comply to specific constraints. Raquel cannot, until update stage  $m$ , rule out that Quassim learnt  $Q_k$  rather than  $Q_0$ , hence we must have

$$P(A|Q_k \cap \lceil r_i \rceil) = q_{i+1}$$

for all  $0 \leq i < m$ . Since these constraints are the same for all  $Q_k \in Q_m$  we can formulate the constraints on that level. Moreover, we can do so for both Raquel and Quassim and for any update stage. We obtain

$$P(A|\mathcal{R}_m \cap \lceil q_i \rceil) = r_{i+1}, \quad (10)$$

$$P(A|Q_m \cap \lceil r_i \rceil) = q_{i+1}, \quad (11)$$

for all  $0 \leq i < m$  and for all  $m \leq n$ . It will be clear that these constraints can also be met in virtue of the Lemma above.

Notice that  $R_0 \subset \mathcal{R}_n$  and  $Q_0 \subset Q_n$ , so that the constraints (8) and (9) are covered by the constraints (10) and (11). Further constraints come from the fact that at stage  $i = m$ , Raquel and Quassim must eliminate elements from the information partition,

$$P(A|\mathcal{R}_m \cap \lceil q_m \rceil) \neq r_{m+1}, \quad (12)$$

$$P(A|Q_m \cap \lceil r_m \rceil) \neq q_{m+1}, \quad (13)$$

for all  $m < n$ . Importantly, because of these inequality constraints, the constraints that  $P(A|\mathcal{R}_m \cap \lceil q_{m-1} \rceil) = r_m$  and  $P(A|Q_m \cap \lceil r_{m-1} \rceil) = q_m$  do not entail anything for the probability assignments within  $\mathcal{R}_m \cap \lceil q_i \rceil$  and  $Q_m \cap \lceil r_i \rceil$  for  $i > m$ .

Now consider what elements of the algebra  $\mathcal{U}$  are involved in satisfying the constraints. We identify the cells  $\mathcal{U}_{jk} = \mathcal{R}_j \cap Q_k$  and observe that the constraints of (10) can be met by adapting the probability assignments of and within all  $\mathcal{U}_{jk}$  for which  $j \geq k$ , while the constraints of (11) involve only the probability assignments of and within  $\mathcal{U}_{jk}$  for which  $j \leq k$ . These constraints are, therefore, pertaining to disjunct, non-overlapping parts of the algebra  $\mathcal{U}$ .

The constraints (12) and (13) do interfere with each other, as they both pertain to the diagonal,  $\mathcal{U}_{jk}$  for  $j = k$ . But they do so in a benign way, as they only require inequality. The constraints can in fact be used to ensure that the constraints imposed above and below the diagonal do not interfere with each other any further. For any row  $\mathcal{R}_{j'}$ , for example, meeting the constraints pertaining to  $\mathcal{U}_{j'k}$  for  $k < j'$  has implications for the probability of  $A$  within  $\cup_{j' \leq k \leq n} \mathcal{U}_{j'k}$ . But because the probabilities in  $\mathcal{U}_{j'j'}$  can be chosen almost entirely freely, the probabilities within the cells  $\mathcal{U}_{j'k}$  for  $k > j'$ , sitting below the diagonal in row  $\mathcal{R}_{j'}$ , are not constrained at all. And the same goes for any specific column  $Q_{k'}$ .

The one remaining way in which the constraints might interfere is through the requirements on the relative sizes, as expressed by Eq. (7) in the Lemma. In the Aumann

structure, this requirement translates to size requirements for both  $P(\ulcorner r_i \urcorner | \mathcal{Q}_m)$  and  $P(\ulcorner q_i \urcorner | \mathcal{R}_m)$  for any  $m$  and any  $i < m$ . The constraints can be met easily by requiring that these conditional probabilities respect the given constraints, substituting, respectively  $r$ 's and  $q$ 's for the  $p$ 's in the Lemma. It is clear that we can always meet these constraints because they all point in the same direction: we simply make sure that  $P(\ulcorner r_i \urcorner \cap \ulcorner q_i \urcorner)$  diminishes rapidly with increasing  $i$ . We conclude that the common prior can accommodate all the constraints.  $\square$

This establishes that the Aumann structure and the common prior can be chosen in such a way that the dynamic approach to common knowledge of posteriors reproduces the process of iterated pooling described by DeGroot. The proof makes good on Aumann's suggestion, cited in the introduction, that his agreement theorem is a theoretical foundation for consensus through iterated pooling.

### 3.3 Extending and refining the result

In this section, we briefly investigate extensions and developments of the above theorem. First we consider the extension to an infinitely long iteration of pooling operations. We then discuss developing the theorem by imposing further constraints on the common prior or by employing the available freedom in the choice of the prior.

The foregoing establishes that a closed-off pooling process can be replicated as an approach to common knowledge by orchestrating the common prior in the right way. But this is not exactly showing that consensus and agreement can be aligned, because strictly speaking the former requires an infinitely long series of operations. One response to this is to point to the limiting properties. Because the closing stage  $n$  may be at any time, and because by increasing  $n$  we can make the discrepancy between  $r_{n-1}$  and  $q_{n-1}$  indefinitely small, we can make the discrepancy between the approach to common knowledge and the consensus formation indefinitely small as well. We have thus shown the identity of consensus and agreement for all practical purposes.<sup>5</sup>

Another response is to draw a parallel between closing off a series of pooling operations, and achieving common knowledge in one update step, as described in [Geanakoplos and Polemarchakis \(1982\)](#). The final update step in the approach to common knowledge is one in which all is revealed. After step  $n$ , neither Raquel nor Quassim has any doubt left as to the posteriors of the other agent. This step differs from preceding steps in the update sequence, in which some variation among the probability assignments within the remaining elements of the information partitions was retained. We might consider the final pooling operation in similar fashion: it encompasses all the considerations that would otherwise play out over infinitely many rounds of pooling. Modeling all these rounds explicitly can be done by expanding the information partitions indefinitely, but in a Bayesian approach to common knowledge such an infinite exchange can also be collapsed onto a single update step.

We now turn to further developments of the theorem. Recall that the theorem leaves considerable freedom in choosing the prior. Consensus through iterated pooling is thus

<sup>5</sup> In various studies on iterated pooling, e.g. [Zollman \(2007\)](#) and [Mayo-Wilson et al. \(2011\)](#), it is customary to apply such a more lenient criterion for consensus.



aligned with a sizeable class of common priors that each lead to the given pooling process. The fact that the representation is not a tight fit invites a more extensive investigation of the systematic relations that obtain between classes of common priors and pooling processes, with which we will make a very modest beginning here.

One suggestion is to constrain the common prior further, by employing the equivalence of pooling and updating that was proved in [Genest and Schervish \(1985\)](#). This equivalence is both more specific and more general than the equivalence of this paper. It only covers a one-shot pooling operation but it is concerned with the level of operations rather than the level of instances of belief change. It entails not merely that the actual probability assignments of Raquel and Quassim match the pooling sequence, but also that the probability assignments of Raquel and Quassim are such that other priors for the proposition of interest would also lead them to the posteriors obtained through a pooling operation.

We will not offer a proof of this claim, but we conjecture that the more strict equality of consensus and agreement is indeed feasible, owing to the same argument that supports our main theorem. The constraints that are needed to establish the more strict equality replace the lenient constraints of the Lemma. But they attach to the very same segments  $\mathcal{U}_{jk}$  of the algebra and so, for the reasons already stated, they do not get in the way of each other. It seems that we can, therefore, construct a more narrow class of common priors that leads to the requisite pooling process, and whose members have the additional feature that a variation of the initial probabilities  $r_1$  and  $q_1$  while leaving the further details of the common prior intact also results in an approach to common knowledge that mimics a pooling process.

Other developments of the representation are concerned with the nature of the pooling operation. For one, until now we have considered linear pooling. But seeing the freedom in choosing the probability assignments that the Lemma still offers, it seems evident that other pooling operations can also be accommodated in a common prior. Further, we might investigate how far we can get in pooling probability assignments over a segment of the algebra rather than a single proposition. Finally, through the representation we have obtained a different characterisation of pooling processes in terms of the corresponding common priors. It will be interesting to investigate the properties of pooling processes by means of their common prior characterisations.

## 4 Discussion

The main representation theorem provides a rationalization of consensus by iterated pooling. For any such consensus formation process starting with non-extremal beliefs, there is a corresponding dynamic agreement model such that, first, the agents in the latter model reach the same agreement as in the iterated pooling scenario and, second, the intermediate opinion dynamics leading to consensus or agreement is the same in both models. In other words, each step in the process of iterated pooling, from first to last, could be taken by fully rational Bayesian agents with common prior beliefs. And the iterative process can be explained by reference to an implicit exchange of information, happening at increasingly higher orders in the belief hierarchy.

It might seem that this result elevates iterated pooling to the same level as Bayesian updating, even though it is conceptually far more parsimonious. But the representation theorem merely shows that any consensus formation process brought about by iterated pooling can be represented in a Bayesian framework. For this representation to work we need to choose a distinct prior probability over the richer algebra of epistemic facts. If we choose another prior probability, we can represent all manner of other processes of opinion exchange. The Bayesian model of agreement is thus far more powerful than the pooling model. It provides a foundation rather than a redescription of consensus through pooling.

One might object that this representation does not provide a satisfactory foundation of consensus formation by iterated pooling because it is too demanding. In iterated pooling, the agents operate on the very simple algebra generated by just one element,  $A$ . In dynamic agreement models, this algebra is much richer. We have seen that it contains propositions regarding the opinions of the others, the opinions of others about the opinions of others, and so on. Furthermore, the agents have sharp beliefs about all elements of this complex algebra. So inasmuch as full opinionation is an unrealistic assumption in probabilistic modeling overall, the situation is aggravated in dynamic agreement models. The argument might be that the rationalization of iterated pooling is only made at the cost of overly strong idealization, which renders it less plausible.

We consider this objection in some detail. We first note that there is certainly nothing irrational, epistemically speaking, in being fully opinionated about a rich algebra of propositions.<sup>6</sup> So the fact that iterated pooling can be rationalized in Bayesian terms is not threatened by observing that full opinionation is not a demand of doxastic rationality.<sup>7</sup> To undermine our representation, the over-demandingness objection might instead appeal to practical considerations, e.g., by arguing that such a system of beliefs about the beliefs of others is computationally too costly for resource-bounded agents, or else to over-idealization. But since we are after an epistemic rationalization, the appeal to practical considerations does not apply. We thus contend that our representation theorem is not too demanding as an *epistemic* rationalization.

Another objection concerns certain peculiarities in the generalization to more than two agents. In the foregoing we have considered only Raquel and Quassim but as is well known, consensus formation and dynamic agreement can both be run with any number of agents. On the side of iterated pooling it is easily seen that this changes little to the nature of the approach to consensus. If we include Simone next to Quassim and have Raquel determine her new probability in response to two agents, we may first combine the opinions of Quassim and Simone using Raquel's weights, and then have Raquel adapt her opinion to the combined opinions of the other two, again by her weights. A parallel reduction to the two agent case can be given on the side of the approach to agreement. However, the probability assignments required for more than two agents have to have certain correlations among the agents built in, if we want the results of updating to match pooling. The Bayesian model thus suggests that pooling

<sup>6</sup> Credal sets may or may not provide a better representation of belief states than sharp probability assignments. However, the case may be, it does not impinge on our discussion, which only gets going once a consensus formation process with sharp probabilities is given.

<sup>7</sup> Witnessing a large body of literature on credal sets in epistemology, see e.g. Bradley (2015).

among more than two agents requires the opinions of the agents to be correlated in a particular way.

The objection might be that such correlations are artifacts of the way in which pooling is represented. But we think not: rather we believe that the Bayesian model offers us a more fine-grained understanding of what we commit to when we engage in iterated pooling. To apply iterated pooling in a situation with three agents, we must assume that the agents have probability assignments that satisfy these constraints, and this is brought out by the Bayesian model. More generally, each consensus formation process can be associated with a class of output-equivalent Bayesian update processes, and these equivalence classes indicate the conditions under which iterated pooling is warranted. Importantly, this viewpoint helps us to answer the criticism that consensus via pooling is too coarse-grained, and that it ignores epistemically relevant aspects of social situations. It may well be that pooling blurs out salient distinctions. But whether these distinctions are unduly neglected when applying iterated pooling depends on the modeler, who will have to judge whether the conditions for applying iterated pooling are met in the case at hand.

Continuing in this more positive mode, we believe that the representation provides a clearer role division for epistemic trust and rational belief change. In pooling models, trust is embodied both by the weights the agents assign to each other and by the belief change mechanism, i.e., the linear pooling operation itself. Learning and trusting are in a sense lumped together. In the dynamic agreement models that we constructed, on the other hand, the two are clearly separated. Learning is covered by the standard Bayesian machinery. Trust, on the other hand, is hard-wired in the common prior and the common knowledge of information partitions, which guarantees a progressive movement towards agreement upon repeatedly learning each others' opinion. In other words, the representation we give is of agents who treat each others' opinion as pieces of evidence like any other, but which are a priori disposed to making concessions towards consensus.

The representation theorem thus shows that the dynamics of beliefs in iterated pooling can be interpreted in terms of trust and, crucially, higher-order beliefs. At each step of the process, the agents not only learn about the posteriors of others, but they also learn what the others have learned about them. This second piece of information needs not affect the agents' beliefs in the factual proposition at hand, that is, in the proposition *A* of the theorem. But it is crucial that it may change their view on what the other person might believe regarding that proposition, thus allowing for further eliminations of elements of the information partition as the process is iterated. So the mutual trust between the agents that is encoded in the prior leads to consensus through a dynamics of higher-order beliefs.

One final worry might be that this interpretation of iterated pooling appears incompatible with the standard one, the one compatible with the classical "peer disagreement problem" (Hartmann et al. 2009; Martini et al. 2013), in which differences of opinion persist even when there is no difference in private information. In such contexts iterated pooling has been suggested as rational arbitration mechanism. If we read sameness of private information in a strict sense, i.e., as having the same information partition and prior, then such cases are obviously ruled out by the assumptions in our representation.

We believe that this does not speak very strongly against our rationalization of iterated pooling, since the purported incompatibility only appears under an extremely strong reading of “sameness of information”. In dynamic agreement models, having the same factual information about the proposition  $A$ , or even about all material facts in the algebra  $\mathcal{A}$ , does not rule out having different higher-order information. The incompatibility with the traditional peer disagreement scenario only appears if we interpret sameness of information as sameness of first and higher-order information. But this puts the bar of being an epistemic peer extremely high. Agents in that scenario are not only equally informed about  $A$ , they are equally informed at all higher-order levels. In that context, it is at the very least questionable whether peer disagreement can still occur.

## 5 Conclusion

Summing up, consensus by iterated pooling can be represented as a dynamic agreement process under common prior. This provides a plausible epistemic rationalization of the former. It does so in a way that highlights potential correlations between the agent’s beliefs in groups larger than two, and that makes a clear distinction between the expression of trust and rational belief change. Iterated pooling and dynamic agreement models should thus not be seen as competing views on how to reach a consensus. In view of our result, they are perfectly compatible, with the latter offering an interpretation of the former.

Bayesian models are of course more general and fine-grained. It is easy to construct a dynamic agreement model to which there corresponds no sequence of iterated pooling, even with changing weights. The example of Sect. 2.3 is a case in point. And the rather large room to maneuver left by the theorem in the construction of the prior suggests that the same iterated pooling sequence can be mapped to rather different epistemic situations. Iterated pooling is thus a special case among the many types of epistemic situations covered by dynamic agreement models, a case that relies on a specific trust that obtains among the agents.

Instead of weighing pros and cons of each side, what the result in this paper suggests is to start traversing the formal bridge between the sides, taking insights and modeling tools along. Among other things, based on the existence of dynamic agreements we might look at generalizations of the pooling model that introduce additional parameters. This may lead to rationalizable consensus formation processes that are better tailored to the practice of consensus formation, without sacrificing too much on the attractive simplicity of consensus through iterated pooling.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6), 1236–1239.
- Bonnay, D., & Cozic, M. (2015). (manuscript). Weighted averaging, Bayesian update, and Jeffrey conditioning.
- Bradley, R. (2007). Taking advantage of difference in opinion. *Episteme: A Journal of Social Epistemology*, 3(3), 141–155.
- Bradley, S. (2015). Imprecise probabilities. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2015 ed.).
- DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69, 118–121.
- French, J. R., Jr. (1956). A formal theory of social power. *Psychological Review*, 63(3), 181–194.
- Geanakoplos, J. D., & Polemarchakis, H. M. (1982). We cannot disagree forever. *Journal of Economic Theory*, 28, 192–200.
- Genest, C., & Schervish, M. J. (1985). Modelling expert judgments for Bayesian updating. *Annals of Statistics*, 13(3), 1198–1212.
- Hartmann, S., Martini, C., & Sprenger, J. (2009). Consensual decision-making among epistemic peers. *Episteme*, 6(02), 110–129.
- Jackson, M. (2008). *Social and economic networks*. Princeton: Princeton University Press, Princeton, NJ.
- Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society: A philosophical and mathematical study*. Berlin: Springer.
- Martini, C., Sprenger, J., & Colyvan, M. (2013). Resolving disagreement through mutual respect. *Erkenntnis*, 78(4), 881–898.
- Mayo-Wilson, C., Zollman, K., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, 78(4), 653–677.
- Steele, K. (2012). Testimony as evidence: More problems for linear pooling. *Journal of Philosophical Logic*, 41, 983–999.
- Stone, M. (1961). The linear opinion pool. *Annals of Mathematical Statistics*, 32, 1339–1342.
- Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), 574–587.